EM Motivation

Let

$$\mathcal{L}(q(z), \theta) = \sum_z q(z) \log \frac{p(x, z | \theta)}{q(z)}$$

and

$$KL(q(z) \| p(z | x, \theta)) = \sum_z q(z) \log \frac{q(z)}{p(z | x, \theta)}$$

where θ is all parameters.

Claim

$$\log p(x | \theta) = \mathcal{L}(q(z), \theta) + KL(q(z) \| p(z | x, \theta))$$

Proof: RHS

$$= \sum_z q(z) \log \frac{p(x, z | \theta)}{p(z | x, \theta)}$$

$$= \sum_z q(z) \log p(x | \theta)$$

$$= \log p(x | \theta)$$

VI derivation (non-intuitive)

Write marginal log-lik. as a sum

our approx

! $$\ln p(x) = \mathcal{L}(q(z)) + KL(q(z) || p(z|x))$$

wher

$$\mathcal{L}(q(z)) = \int q(z) \ln \frac{p(x,z)}{q(z)} dz$$

$$KL(q(z)||p(z|x)) = \int q(z) \ln \frac{q(z)}{p(z|x)} dz$$

Verification (as for EM)

$$\mathcal{L} + KL$$

$$= \int q(z) \ln \frac{p(x,z)}{p(z|x)} dz$$

$$= \int q(z) \ln p(x) dz$$

$$= \ln p(x)$$

We minimize $KL(q(z)||p(z|x))$ by maximizing $\mathcal{L}(q(z))$ (easier, often possible).

VI inference

We minimize $KL(q(z)\|p(z|x))$ by maximizing $\mathcal{L}(q(z))$.

Max $\mathcal{L}(q(z))$

Assume $q(z)=\prod_i q_i(z_i)$ $q_i(z_i)$ written $q(z_i)$

We iteratively maximize $\mathcal{L}(q(z))$ by max. w.r.t. $q(z_j)$ (with varying j)

$$\mathcal{L}\left(\prod_i q(z_i)\right)$$

$$= \int_z \prod_i q(z_i) \log \frac{p(x,z)}{\prod_e q(z_e)} \, dz$$

$$= \int \prod_i q(z_i) \left[\log p(x,z) - \sum_e \log q(z_e)\right] dz$$

$$= \int q(z_j) \left[\int \prod_{i\neq j} q(z_i) \log p(x,z) \, dz_{-j}\right] dz_j$$

$$- \sum_e \int \prod_i (z_i) \log q(z_e) \, dz = *$$

Let $\log \hat{p}(x, z_j) = E_{\prod_{i \neq j} q(z_i)} [\log p(x, z)]$

$$* \doteq \int q(z_j) \log \hat{p}(x, z_j)\, dz_j$$

$$- \int q(z_j) \log q(z_j)\, dz_j$$

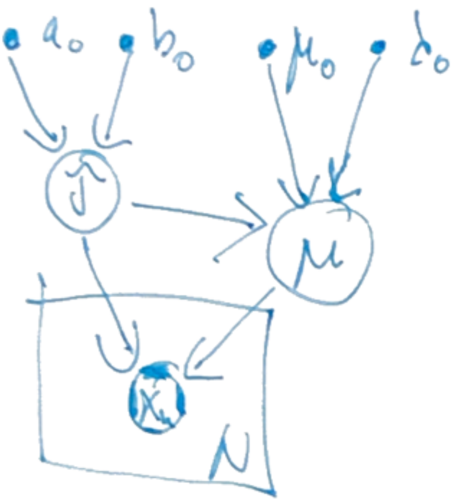
$$= \int q(z_j) \log \frac{\hat{p}(x, z_j)}{q(z_j)}\, dz_j = -KL(q(z_j) \| \hat{p}(x, z_j))$$

So $\mathcal{L}(\prod_i q(z_i))$ max by $q(z_j) = \hat{p}(x, z_j)$, i.e.,

by $\log q(z_j) = \log \hat{p}(x, z_j) = E_{\prod_{i \neq j} q(z_i)} [\log p(x, z)]$

This gives the update equation for $q(z_j)$

10.1.3 Univariate Normal



$$\tau \mid a_0, b_0 \sim Ga(\tau \mid a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)} \tau^{a_0 - 1} e^{-\tau b}$$

$$\mu \mid \mu_0, \lambda_0, \tau \sim \mathcal{N}(\mu \mid \mu_0, (\tau\lambda_0)^{-1})$$

$$= \sqrt{\frac{\tau\lambda_0}{2\pi}} \, e^{-\frac{\tau\lambda_0}{2}(\mu - \mu_0)^2}$$

$$x_n \mid \mu, \tau \sim \mathcal{N}(x_n \mid \mu, \tau^{-1}) = \sqrt{\frac{\tau}{2\pi}} \, e^{-\frac{\tau}{2}(x_n - \mu)^2}$$

Data $\mathcal{D} = \{x_1, \dots, x_N\}$

$$p(\mathcal{D} \mid \mu, \lambda) = \left(\frac{\tau}{2\pi}\right)^{\frac{N}{2}} e^{-\frac{\tau}{2}\sum_{n=1}^{N}(x_n - \mu)^2}$$

and

$$p(\mathcal{D}, \mu, \tau) = p(\mathcal{D} \mid \mu, \tau)\, p(\mu \mid \tau)\, p(\tau)$$

Update equation

$\log q^*(\mu)$

$\stackrel{+}{=} \mathbb{E}_\tau[\log p(D|\mu,\tau) + \log p(\mu|\tau)]$

$\stackrel{+}{=} -\frac{\mathbb{E}_\tau[\tau]}{2}\left(\sum_n (x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right)$

Completing the square gives

$$q^*(\mu) = N(\mu \mid \mu_*, \lambda_*^{-1}$$

where

$$\lambda_* = \mathbb{E}_\tau[\tau](N + \lambda_0)$$

$$\mu_* = \frac{\left(\sum_n x_n\right) + \lambda_0\mu_0}{N + \lambda_0}$$

$$\log q^*(\tau)$$

$$\overset{+}{=} E_\mu \left[\log p(D \mid \mu, \tau) + \log p(\mu \mid \tau) \right] + \log p(\tau)$$

$$\overset{+}{=} -\frac{\tau}{2} E_\mu \left[\sum_n (x_n - \mu)^2 + \lambda_0 (\mu - \mu_0) \right]$$

$$+ \frac{N+1}{2} \log \tau + (a_0 - 1) \log \tau - b_0 \tau$$

So $q^*(\tau) = \mathrm{Ga}(\tau \mid a_*, b_*)$

where

$$a_* = a_0 + \frac{N+1}{2}$$

$$b_* = b_0 + \frac{1}{2} E_\mu \left[\sum_n (x_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right]$$

The expectation can be computed,
since $E[\mu]$ is known $V[\mu]$ is known

and $E[\mu^2] = V[\mu] + E[\mu]^2$